Metabarcoding and Metagenomics 2: 1–12
DOI 10.3897/mbmg.2.30457

Research Article

Monitoring riverine fish communities through eDNA metabarcoding: determining optimal sampling strategies along an altitudinal and biodiversity gradient

Jonas Bylemans[1,2], **Dianne M. Gleeson**[1,2], **Mark Lintermans**[1], **Christopher M. Hardy**[2,3], **Matthew Beitzel**[4], **Dean M. Gilligan**[2,5], **Elise M. Furlan**[1,2]

1 *Institute for Applied Ecology, University of Canberra, Canberra, ACT 2617, Australia.*
2 *Invasive Animals CRC, University of Canberra, Canberra, ACT 2617, Australia.*
3 *CSIRO Land and Water, GPO Box 1700, Canberra, ACT 2601, Australia.*
4 *Conservation Research ACT Government, GPO Box 158, Canberra, ACT 2601, Australia.*
5 *NSW Department of Primary Industries - Fisheries, PO Box 17, Batemans Bay, NSW 2536, Australia.*

Corresponding author: *Jonas Bylemans* (Jonas.Bylemans@canberra.edu.au)

Academic editor: *D. Steinke* | **Received** 10 October 2018 | **Accepted** 16 December 2018 | **Published** 31 December 2018

Abstract

Monitoring aquatic biodiversity through DNA extracted from environmental samples (eDNA) combined with high-throughput sequencing, commonly referred to as eDNA metabarcoding, is increasing in popularity within the scientific community. However, sampling strategies, laboratory protocols and analytical pipelines can influence the results of eDNA metabarcoding surveys. While the impact of laboratory protocols and analytical pipelines have been extensively studied, the importance of sampling strategies on eDNA metabarcoding surveys has not received the same attention. To avoid underestimating local biodiversity, adequate sampling strategies (i.e. sampling intensity and spatial sampling replication) need to be implemented. This study evaluated the impact of sampling strategies along an altitudinal and biodiversity gradient in the upper section of the Murrumbidgee River (Murray-Darling Basin, Australia). An eDNA metabarcoding survey was used to determine the local fish biodiversity and evaluate the influence of sampling intensity and spatial sampling replication on the biodiversity estimates. The results show that optimal eDNA sampling strategies varied between sites and indicate that river morphology, species richness and species abundance affect the optimal sampling intensity and spatial sampling replication needed to accurately assess the fish biodiversity. While the generality of the patterns will need to be confirmed through future studies, these findings provide a basis to guide future eDNA metabarcoding surveys in river systems.

Key Words

Environmental DNA, eDNA, Metabarcoding, Sampling, Fish

Introduction

Robust methods for monitoring species biodiversity are the fundamental basis for ecological research and environmental management. High-Throughput Sequencing (HTS) of PCR amplicons derived from environmental DNA (eDNA), commonly referred to as eDNA metabarcoding, is becoming an increasingly popular tool for such monitoring surveys (Creer et al. 2016, Bohmann et al. 2014, Taberlet et al. 2012). However, like any method, eDNA metabarcoding can suffer from both false positive and false negative detections (Shelton et al. 2016, Ficetola et al. 2015, Shaw et al. 2016). Consequently, before this technology can be used as a standard monitoring tool, sampling protocols need to be evaluated and optimal sampling strategies developed.

The popularity of eDNA-based monitoring of aquatic biodiversity has increased dramatically since the first published study (Ficetola et al. 2008, Rees et al. 2014, Deiner et al. 2017). Early studies primarily focussed on

Copyright *Jonas Bylemans et al.* This is an open access article distributed under the terms of the Creative Commons Attribution License (CC BY 4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

single species detections (Ficetola et al. 2008, Jerde et al. 2011, Goldberg et al. 2011), but eDNA metabarcoding is increasingly being used to characterize entire species communities (Valentini et al. 2016, Hänfling et al. 2016, Bista et al. 2017, Blackman et al. 2017). Within the current literature, there has been a strong focus on the optimization of laboratory methods and bioinformatics analyses to reduce potential biases (Ficetola et al. 2016). Protocols used to capture and extract eDNA, amplify the barcoding region of interest, and construct HTS libraries are known to influence biodiversity estimates (Kanagawa 2003, Elbrecht and Leese 2017, Deiner et al. 2015, O'Donnell et al. 2016, Schnell et al. 2015). The impact of such biases on eDNA metabarcoding surveys can generally be reduced by increasing the replication in laboratory protocols (Ficetola et al. 2015, Alberdi et al. 2017, Sato et al. 2017). Although replication in laboratory procedures has received considerable interest, the importance of sampling strategies (i.e. sampling intensity and spatial sampling replication) is poorly understood (Dickie et al. 2018).

Aquatic biodiversity estimates obtained from conventional monitoring surveys are known to be influenced by the type of sampling method used (Kennard et al. 2006, Porreca et al. 2013, Bower et al. 2014), the sampling season (Fischer and Quist 2014, Jurajda et al. 2009) and the sampling effort (Reid and Haxton 2017, Pritt and Frimpong 2014, Ebner et al. 2008). In order to obtain robust biodiversity estimates, a thorough understanding of the potential biases of survey methods is essential (Gotelli and Colwell 2001, Schmidt 2005). Seven eDNA metabarcoding studies to date have evaluated the effect of sampling intensity on biodiversity estimates. Three studies have focussed on the freshwater fish biodiversity of lentic systems and have confirmed that sampling intensity is important to accurately characterize the entire fish community (Evans et al. 2017, Sato et al. 2017, Hänfling et al. 2016). However, the importance of spatial sampling replication is unclear as Evans et al. (2017) reported a low spatial heterogeneity between sampling replicates collected from a single lake while other studies found clear spatial patterns (Sato et al. 2017, Hänfling et al. 2016). Four studies to date have varyingly addressed the impact of sampling intensity for eDNA metabarcoding surveys in lotic systems (Li et al. 2018, Olds et al. 2016, Shaw et al. 2016, Pont et al. 2018). While Shaw et al. (2016) found that an increased sampling intensity will increase the observed species richness, the remaining studies report only a moderate increase in species richness with an increased sampling effort (Olds et al. 2016, Pont et al. 2018, Li et al. 2018). However, two of the previously published studies focussed on relatively small streams (i.e. Olds et al. 2016, Li et al. 2018) and for all studies the sampling intensity (e.g. Li et al. 2018) or the spatial sampling replication (e.g. Pont et al. 2018) was limited.

It can be expected that in freshwater lotic systems, the importance of eDNA sampling strategies (i.e. sampling intensity and spatial sampling replication) will increase in higher order streams. Firstly, it is well known that in lotic systems the fish species richness changes along an altitudinal gradient with higher order streams (i.e. lower altitude) supporting a higher species richness due to an increased habitat size and/or habitat diversity (Tejerina-Garro et al. 2005, Unmack 2001). As local species richness increases it can be expected that the probability of missing a species increases and thus a higher sampling intensity may be needed to accurately assess the species biodiversity (Ficetola et al. 2015, Yamamoto et al. 2017). Furthermore, the increased species richness and habitat diversity at low altitude sites will increase competitive interactions and niche differentiation (Tejerina-Garro et al. 2005). This in turn will increase the spatial structuring of the community and may result in a more heterogeneous distribution of eDNA (Stoeckle et al. 2017, Port et al. 2016). River channel characteristics such as river width, depth and water flow are also likely to influence the spatial distribution of eDNA. As high-altitude streams tend to be relatively narrow, shallow and fast flowing it can be expected that eDNA is more homogeneously distributed in these systems. In contrast, the wider, deeper and slower flowing water in low-altitude sites are likely to result in a more heterogenous eDNA distribution. Finally, season fluctuations in flow, temperature etc. will also influence the distribution of eDNA but are not the primary focus of the current study (Bista et al. 2017, Jane et al. 2015).

Here, we hypothesize that both the species richness and the river channel morphology (e.g. river width, depth and water flow) will influence the optimal sampling strategies (i.e. sampling intensity and spatial sampling replication) for eDNA metabarcoding surveys. To test this, we used eDNA metabarcoding to assess the fish biodiversity at five sites along an altitudinal and biodiversity gradient within a single catchment. A systematic sampling approach was used at each site to evaluate the impact of sampling strategies on the observed species richness derived from eDNA metabarcoding data.

Materials and methods

Sampling sites

This study examines the fish biodiversity within a single catchment of the Murray-Darling Basin (MDB) (Australia). River morphology, abiotic water conditions and fish biodiversity vary along an altitudinal gradient in the MDB (Lintermans 2007). High altitude river systems are characterized by rocky substrates, high flow rates, clear waters, and low fish biodiversity. Low altitude river systems generally have a soft bottom substrate, are slow-flowing, turbid and contain higher fish biodiversity. Environmental DNA sampling was conducted from the upper section of the Murrumbidgee River (MR) catchment at five sites within the main channel. Sampling sites were selected based on their altitude and expected species diversity (Figure 1) (Lintermans 2002). Sampling sites ranged from high (1,303 m above sea-level at

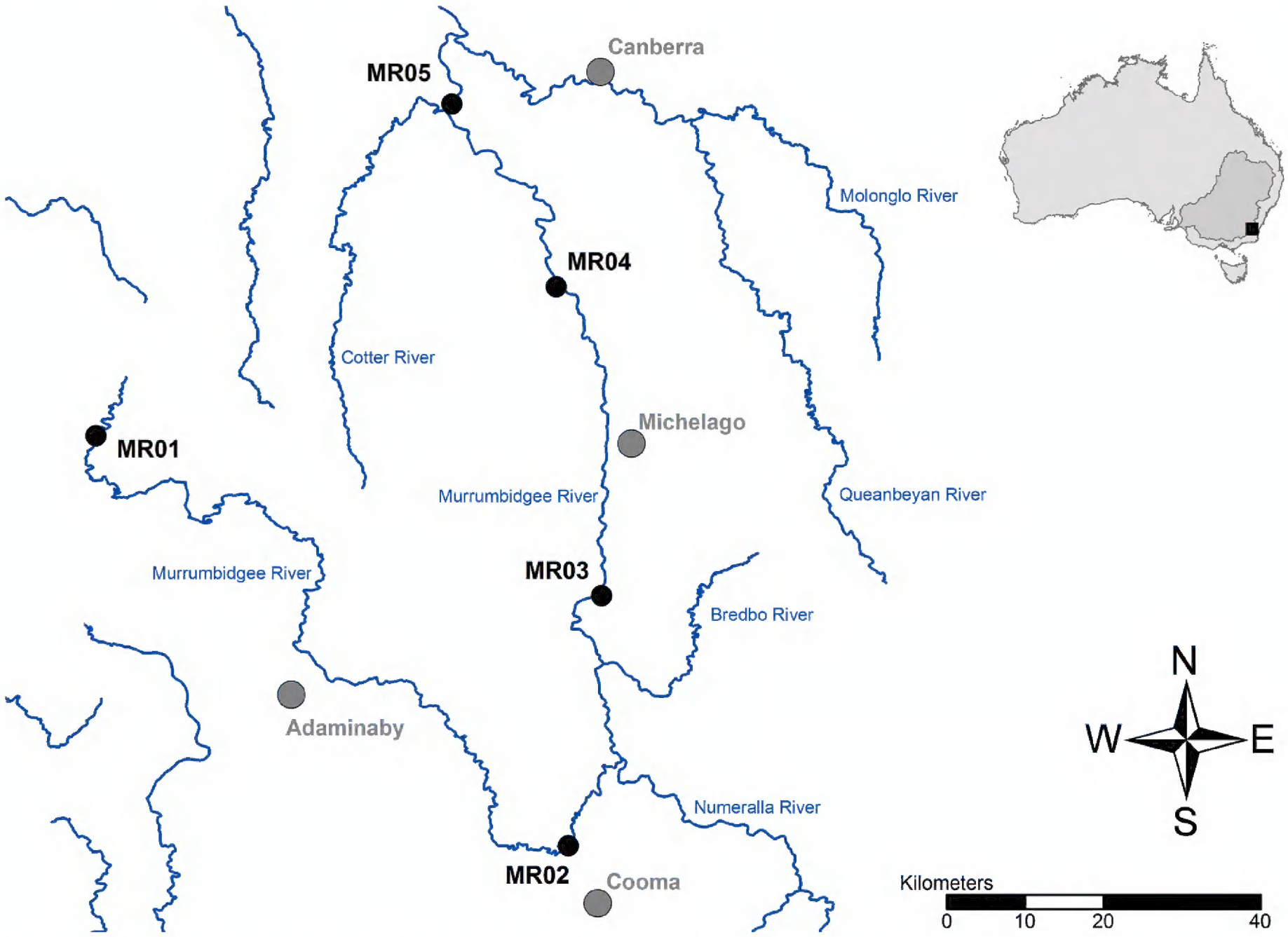

Figure 1. Map of all sampling sites within the main channel of the Murrumbidgee River (MR). Sampling sites are numbered from upstream (MR01) to downstream (MR05).

MR01) to low (520 m above sea-level at MR05) altitude with a minimal distance of *ca.* 35 km between sampling sites (measured along the river channel). The average river width at each sampling site was 11.75 m (MR01), 33.75 m (MR02), 58.25 m (MR03), 52.00 m (MR04) and 52.25 m (MR05).

Monitoring of the fish biodiversity

To obtain a measure of expected fish biodiversity, which is required to evaluate the overall performance of the eDNA metabarcoding survey, an expert opinion survey was conducted. Additionally, conventional boat electrofishing data was available for the three most downstream sites. A detailed description of the expert opinion survey data and the electrofishing data is available in the supplementary materials (Suppl. material 1).

Environmental DNA samples (i.e. twelve 2 L water samples) were collected from each site between the 3rd and 8th of November 2016 (only eleven samples were available from the MR01 site as one of the sampling bottles broke). Potential contaminating DNA was removed from sampling equipment prior to collecting water samples, using a 20% bleach solution and thoroughly rinsing with UV-sterilized tap water. One blank field control (BFC) was included for each site and consisted of a 2 L sampling bottle filled with UV-sterilized water which was opened on site, closed and submerged in the water. At each site, samples were collected over four transects across the river width spanning a 100 m river section (i.e. *ca.* 33 m distance between transects). At the MR01 site samples were collected by wading in the river while for the remaining sites all samples were collected from a canoe. Along each transect surface water samples were collected from both the left and right river banks (i.e. within 1m of the river bank) and the mid-channel. Samples were stored on ice, transported to the University of Canberra (ACT, Australia) and eDNA was captured within 12 h using a 1.2 µm glass fibre filter (Sartorius, Göttingen, Germany). Filtering equipment was cleaned between samples as described above and negative equipment controls (NEC) were obtained by filtering 500 mL of UV-sterilized tap water prior to processing water samples. Filters were stored at -20 °C until eDNA extractions were performed at the trace DNA laboratory (University of Canberra) using the PowerWater DNA Extraction Kit (MoBio Laboratories, Carlsbad, USA). For each sampling site one BFC and two NEC were included in the batch DNA extractions to monitor contamination and all eDNA extracts were stored at -20 °C for further analyses.

PCR amplification and the construction of the HTS libraries was done using the MiFish-U universal fish

primers (Miya et al. 2015). These primers were selected based on the results of an *in silico* primer evaluation of the local fish diversity (Bylemans et al. 2018). Real-Time PCR reactions were run in triplicate for each sample with 25 µL individual reactions consisting of 0.20 µL of AmpliTaq Gold DNA Polymerase (5 U/µL; Applied Biosystems, Foster City, CA, USA), 2.50 µL of GeneAmp 10× Gold Buffer (Applied Biosystems), 2.00 µL of $MgCl_2$ (25 mmol/L; Applied Biosystems), 0.65 µL of GeneAmp dNTP Blend (10 mmol/L; Applied Biosystems), 0.20 µL UltraPure BSA (50 mg/ml; Invitrogen), 0.60 SYBR Green I Nucleic Acid Gel Stain (5X; Invitrogen), 1.00 µL of each primer (10 µmol/L), and 4.00 µL of template DNA and DEPC-treated water (Invitrogen). All PCRs were run on a Bio-Rad CFX96 Real-Time PCR System (Bio-Rad Laboratories, Hercules, USA) with PCR thermal cycling conditions consisting of 5 min at 95 °C; 45 3-step cycles of 30 s at 95 °C, 30 s at 61.5 °C, and 1 min at 72 °C; a final extension of 10 min at 72 °C and a melting curve with a stepwise increase of 0.1 °C/5 s from 60 to 95 °C. When negative control samples tested positive for fish eDNA, they were included in the HTS library preparation protocol. HTS libraries were constructed using a one-step Real-Time PCR amplification with fusion tagged primers (FTP) and the reaction and cycling conditions described above. Forward FTP consisted of the P5 sequencing adaptor, a custom forward sequencing primer, a 7 bp Multiplex Identification (MID-) tag and the MiFish-U forward primer. Reverse FTP contained the P7 sequencing adaptor, a 7 bp MID-tag and the MiFish-U reverse primer. MID-tags were generated using the EDITTAG scripts and unique combinations of forward and reverse MID-tags were used to label amplicons from individual samples (Faircloth and Glenn 2012). For each sample, PCR reactions were performed in triplicate and amplicon libraries of 9 to 10 samples were pooled based on the average Ct-value. Pooled libraries were cleaned using Agencourt AMPure XP Beads (Beckman Coulter, Brea, USA) in a 1.2 volume ratio. Gel-electrophoresis (i.e. 2% agarose gel with a run time of 30 min at 120 V) was used to confirm the presence of a single amplicon and amplicons from each library pool were combined into a single super pool based on the band intensities. A total of 259 uniquely labelled amplicon libraries were combined in the final super pool that consisted of 60 libraries from the current study (i.e. 59 libraries from eDNA samples and 1 library from a BFC) and 199 libraries from other research projects. Unidirectional sequencing was conducted at the Ramaciotti Centre for Genomics (University of New South Wales) using the MiSeq v2 1x300bp sequencing kit.

Bioinformatics

Trimmomatic v.0.36 was used to trim technical sequences (i.e. sequencing adaptors and primers) from the sequencing reads (Bolger et al. 2014). Simultaneously low-quality bases (i.e. Q-score below 3) at the end of the sequencing reads were removed and a sliding window of 4-bases was used to trim reads when the average quality per base was below 15. Further bioinformatics filtering of the sequence reads followed the general workflow described in De Barba et al. (2014). The OBITOOLS scripts were used to assign sequence reads to their respective samples and remove sequences less than 150 bp in length and with an occurrence below 100 (Boyer et al. 2016). The threshold for removing low abundance sequences was determined experimentally so that all Actinopterygii sequences were removed from negative control samples. Sequences arising from PCR and sequencing errors were removed using the OBICLEAN and OBIGREP scripts and taxonomic information was assigned to the sequence records using the ECOTAG script. The reference database used for taxonomic assignments was built using the standard vertebrate sequences from the EMBL data repository (release 132) and custom 12S sequences of all Actinopterygii species in the MDB (Bylemans et al. 2018).

Further filtering and analyses of the metabarcoding data was done using the packages tidyverse (Wickham 2016), INEXT (Hsieh et al. 2016) and vegan (Oksanen et al. 2007) in R version 3.4.1 (Suppl. material 2) (R Development Core Team 2010). The extra quality check revealed that some reads are likely to represent PCR and/or sequencing errors. Within the MR01 site *Salmo trutta* reads were highly abundant while the highest taxonomic assignment of a small proportion of the reads was the *Salmo* genus. The latter reads are likely to represent PCR and/or sequencing variants of *S. trutta* reads and were thus reassigned to *S. trutta*. A single sample of the MR05 site also returned reads assigned to *Salmo salar* and no other Salmonid species were detecting in this sample. As *S. salar* is unlikely to be present at this site these sequences likely represent contaminant DNA and were thus removed. As this site is a popular recreation spot one possible explanation could be the contamination by recreational users. Prior to further data analyses, species names were modified to ensure consistency between the expert opinion, electrofishing and metabarcoding data. As the barcoding region used in this study is unable to distinguish between species in the *Galaxias* genus all species were grouped into a single genus level taxonomy (Bylemans et al. 2018). Finally, all *Hypseleotris* species were renamed to *H. klunzingeri* as it is the only species of this genus to occur in the upper section of the Murrumbidgee River (Lintermans 2002).

Data analyses

The overall performance of the eDNA metabarcoding survey was evaluated by comparing the observed species richness with the estimated species richness obtained from both the expert opinion survey and the electrofishing data (Suppl. material 1).

To determine the impact of sampling intensity and spatial sampling replication the eDNA metabarcoding data was used to construct species accumulation curves (SAC), the total estimates species richness (S_{est}) and the number of samples required to detect 95% of S_{est} was determined. The data obtained from the MR01 site was

excluded from the analyses as no variation in species detections were observed between samples. For the remaining sites, two sampling strategies were evaluated with the analyses using all available sample replicates or only those samples collected from the river-banks (i.e. LB and RB). The community data was first converted to presence/absence data and species accumulation curves (SAC) were constructed for each site and sampling strategy using the iNEXT function (R package iNEXT) (Hsieh et al. 2016). Additionally, the data was used to determine S_{est} (i.e. Chao2 estimates) and evaluate the number of samples required to detect 95% of S_{est}.

As sequencing depth (i.e. the number of reads per sample) could also influence the detection of species from metabarcoding analyses, the metabarcoding data with absolute read counts was used to determine the impact of sequencing depth. A custom R script relying on the R package vegan was used to first rarefy the community data to simulate different sequencing depths (i.e. 10,000; 20,000 and 40,000 reads per sample) (Oksanen et al. 2007). During this rarefaction both reads assigned to Actinopterygii species and reads discarded during the bioinformatics filtering process were considered (Suppl. material 1). Next, non-Actinopterygii reads were discarded and Actinopterygii reads with counts below 100 (i.e. the threshold used in the bioinformatics filtering process) were removed. The data was transformed to presence/absence data before using the iNEXT function to construct SAC. Finally, rank abundance curves (RAC) were also constructed to express the relative read abundance for each species at each sampling site.

A permutational multivariate analysis of variance was used for a more in-depth evaluation of the impact of spatial sampling replication. Metabarcoding results were transformed to presence/absence data and community level differences between sampling transects and locations (i.e. left bank, mid-channel and right bank) were assessed using the ADONIS function within the R package vegan (Oksanen et al. 2007). Sampling transects and locations were set as independent variables and permutations were constrained within sites. When significant effects were observed, the SIMPER function was used to evaluate the average contribution of each species to the overall dissimilarity.

Results

Bioinformatics

After assigning the sequence reads to their respective samples an average of 31,365 sequence reads were obtained per sample with a minimum and maximum sequencing depth of 6,907 and 52,742 reads per sample, respectively. The overall quality of run was high (PhredQ30 score ≥ 91.17). The effect of the bioinformatics filtering processes on the number of sequencing reads for each sample is shown in the supplementary materials (Suppl. material 1).

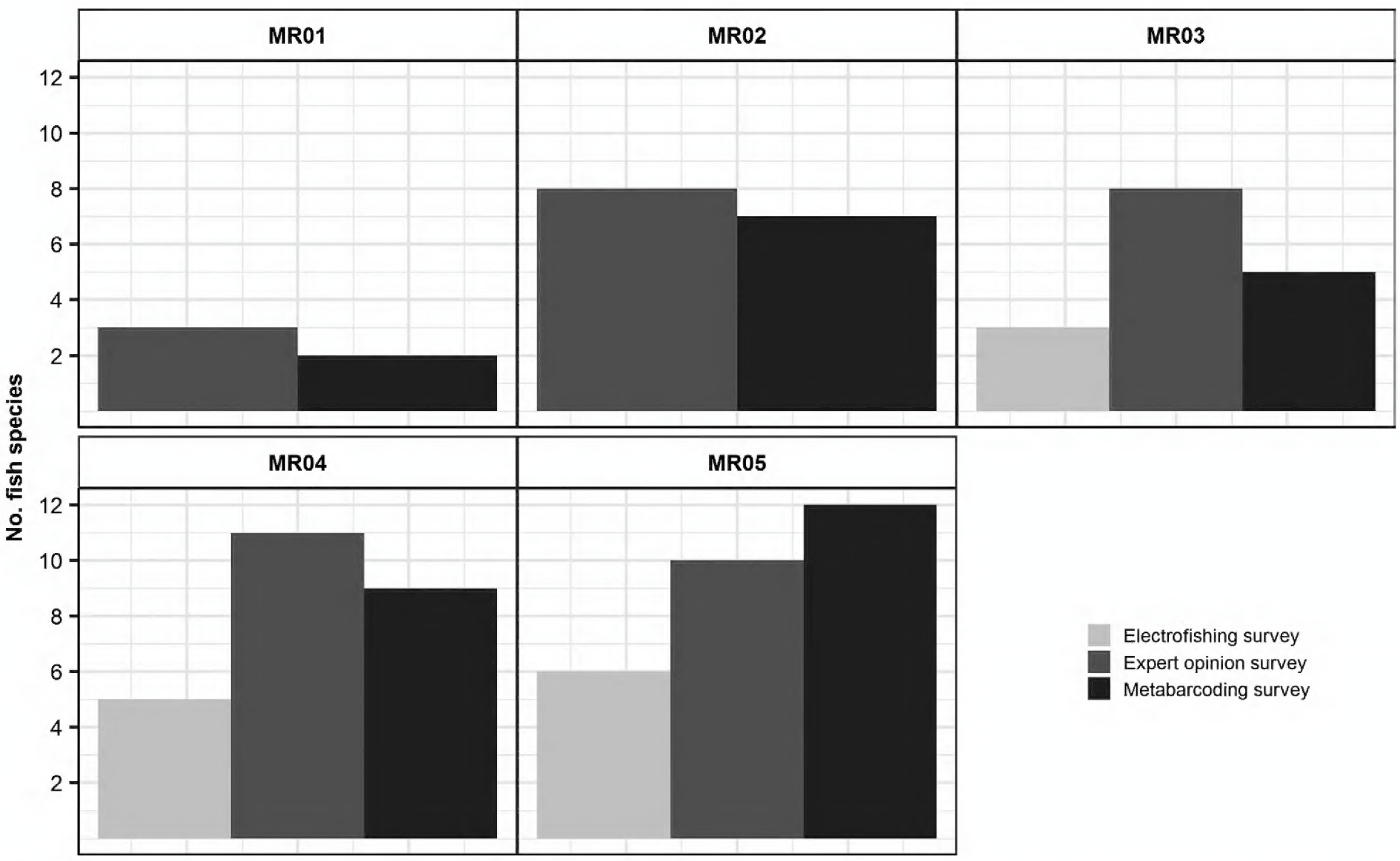

Figure 2. The species richness for each sampling site within the Murrumbidgee River obtained from the standard electrofishing surveys, the expert opinion survey and the eDNA metabarcoding survey.

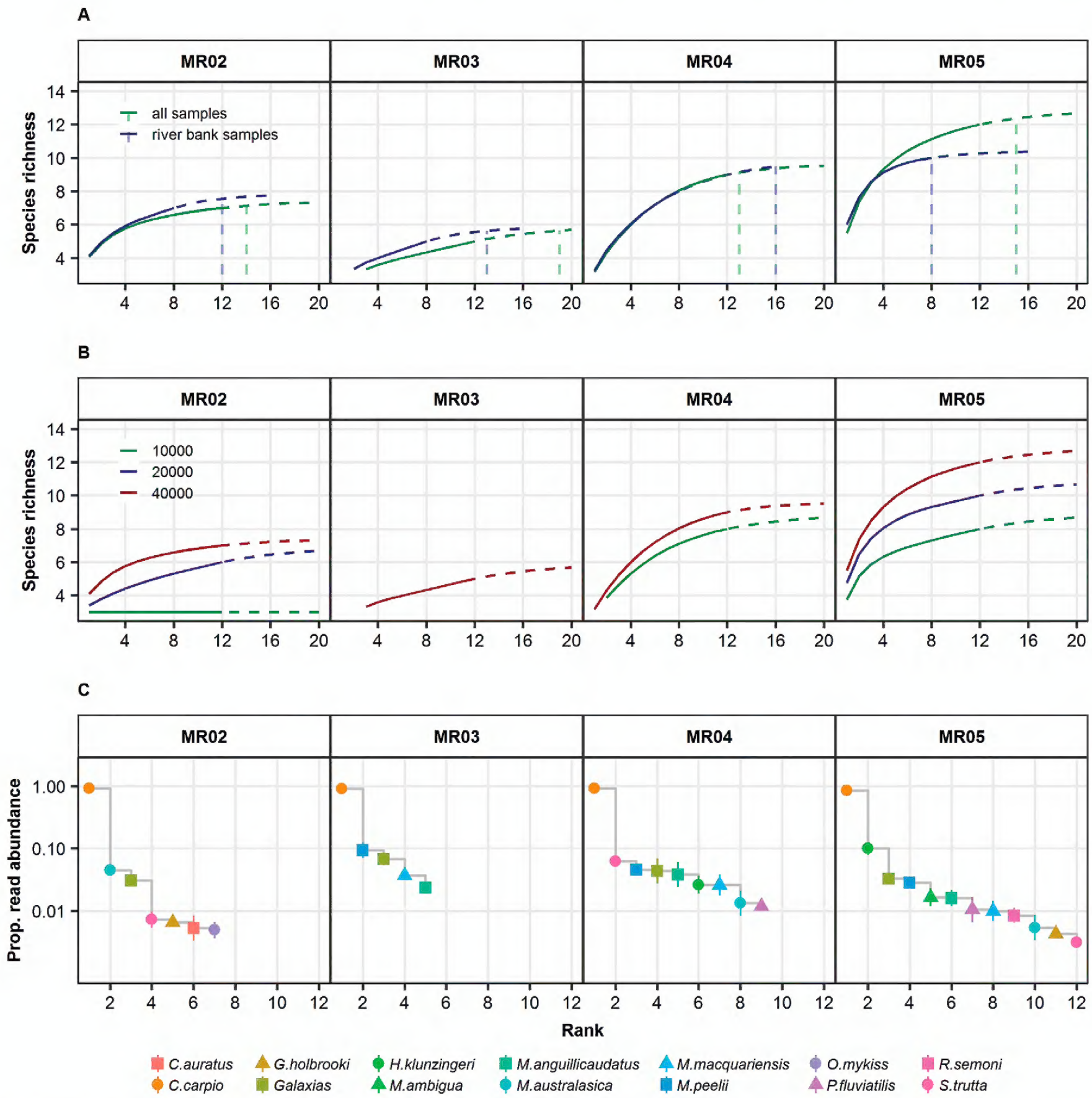

Figure 3. Species accumulation curves (SAC) (**A–B**) and rank abundance curves (RAC) (**C**) for the four most downstream sampling sites in the Murrumbidgee River. SAC were constructed to compare two different sampling strategies (**A**) (i.e. using all available samples and only samples collected from the river-banks) and different sequencing depths (**B**). The vertical dashed lines in panel **A** show the number of samples needed to detect ≥ 95% of the estimated species richness (i.e. based on the Chao2 estimates) for each site and sampling strategy.

Performance evaluation of eDNA metabarcoding

The estimated/observed species richness for each sampling site and each survey method (i.e. expert opinion survey, electrofishing data and eDNA metabarcoding survey) is given in Figure 2. The results show the expected pattern of increasing richness with decreasing altitude. The total species richness of the two most upstream sites were highly similar for both the expert opinion and the eDNA metabarcoding surveys. However, the expert opinion data suggested that for each site one additional species was present but not detected during the metabarcoding survey (*Galaxias* sp. and *Maccullochella macquariensis* are believed to be present at the MR01 and MR02 sampling site respectively) (Suppl. material 1). The data from the three most downstream sites shows that the species richness obtained from the electrofishing data is consistently lower than the species richness derived from both the expert opinion and metabarcoding survey (Figure 2). Only on two occasions were species detected by the electrofishing surveys and not by the eDNA metabarcoding survey (i.e. *Macquaria australasica* at the MR03 site and *Macquaria ambigua* at the MR04 site) (Suppl. material 1). Contrastingly, three, five, and six species were only detected by the eDNA metabarcoding survey at the MR03, MR04 and MR05 site respectively (Suppl. material 1).

Effect of sampling strategies for eDNA metabarcoding surveys

The SAC constructed for each site using all samples and the river-bank samples only show that the number of samples needed to characterize the species community generally increases with decreasing altitude and an increase in species richness (Figure 3A). For both the MR02 and MR03 sites sampling the river-banks only is an appropriate strategy with 12 and 13 samples being sufficient to detect ≥ 95% of S_{est}, respectively (Figure 3A). The results of the MR04 site reveal that approximately 16 samples collected from both river-banks would be adequate to detect 95% of S_{est}. However, increasing the spatial sampling replication reduced the number of samples required to characterize the species richness (i.e. the collection of approximately 13 samples from both river-banks and the mid-channel allows for the detection of 95% of S_{est}) (Figure 3A). Finally, at the most downstream site (i.e. MR05) the results show that only sampling the river-banks will result in an underestimation of the true biodiversity and collecting approximately 15 samples from both river-banks and mid-channel is sufficient to detect ≥ 95% of S_{est} (Figure 3A).

The results from the SAC derived from the data with different simulated sequencing depths show that, for most sites, increasing sequencing depth only moderately improves species detections (Figure 3B). While there is no obvious effect of species diversity on the impact of sequencing depth, when considering both the SAC and the RAC the results show that sequencing depth only has an impact for the MR02 and MR05 sites (Figure 3B, C). The results of the RAC show that for these sites some species have a very low proportional abundance of sequence reads (i.e. below 1% of the total sequence reads per samples) and these species are likely to be missed when using a sequencing depth of 10,000 reads per sample.

The results of the analysis of variance of the community data reveal the need for spatial sampling replication at low altitude sites with a high species richness and is consistent with the results of the SAC. The analysis of variance of the community data showed that sampling location had a significant effect on the community data ($R^2 = 0.01931$, p-value = 0.043) while sampling transect did not ($R^2 = 0.00618$, p-value = 0.876). The results show that the overall community dissimilarity between sampling locations is relatively low for the MR02 site compared to the three most downstream sites (Figure 4A). No clear patterns were observed in the overall community dissimilarity for the MR03 and MR04 sampling sites. The results from the MR05 sampling site, however, show that the community dissimilarity is higher between the mid-channel sampling location and the river-bank sampling locations (i.e. LB and RB) (Figure 4A). In all three downstream sampling locations the detections of *Galaxias* sp. appear to be relatively important in determining the community dissimilarity between mid-channel samples and river-bank samples (Figure 4B). *Galaxias* sp. detections were overall higher for the river-bank samples collected at the MR03 and MR05 sites, while the reverse pattern was found for the MR04 site (Suppl. material 1). The most noticeable community differences between sampling locations were found in the MR05 sampling site. Here four fish species (i.e. *Gambusia holbrooki*, *Galaxias* sp., *Maccullochella peelii* and *S. trutta*) contribute to the community dissimilarity between the mid-channel sampling location and the river-bank sampling locations but did not contribute to the dissimilarity between the two river-bank locations (Figure 4B). While *G. holbrooki* and *S. trutta* were exclusively detected in the mid-channel samples collected at the MR05 site; *M. macquariensis* and *Perca fluviatilis* detections were only found in the river-bank samples (Suppl. material 1).

Discussion

Data obtained here shows that optimal sampling strategies vary along altitudinal and biodiversity gradients. These findings are unlikely to be exclusive to our study system as the increase in fish species richness with decreasing altitude is a global pattern (Platts 1979, Oberdorff et al. 1993, Smith and Kraft 2005). The results of this study show that consideration of river morphology and the local fish community (i.e. expected species richness and evenness) can help guide the sampling design.

Comparing the observed species richness from both electrofishing data and eDNA metabarcoding data clearly shows that standard electrofishing surveys are likely to underestimate the true species richness. Overall, the results show that eDNA metabarcoding detects approximately double the species richness compared to the electrofishing data (Figure 2). Electrofishing is known to be biased and this could explain the nine failed detection in the electrofishing survey data while both the experts and the metabarcoding analyses indicated the species presence (Suppl. material 1). Small-bodied fish are often underrepresented in boat-electrofishing surveys which may explain the failed detections for *G. holbrooki*, *H. klunzingeri* and *Misgurnus* sp. (Kennard et al. 2006, Ruetz et al. 2007). Species-specific biases of conventional sampling methods are also well documented (e.g. Clavero et al. 2006, Bies et al. 2016); with boat electrofishing known to result in a high rate of false negative detections for *Macquaria australasica* (Lintermans 2016). Cryptic, benthic, fossorial species such as *Misgurnus anguillicaudatus* are also known to be difficult to capture by electrofishing (Hinlo et al. 2018, Sigsgaard et al. 2015), particularly in large stream habitats with elevated turbidity (Lintermans unpublished). The electrofishing data also failed to detect trout cod (*Macquaria macquariensis*) at three sites where both the experts and the metabarcoding survey indicated its presence, which could indicate that the current survey methods are inadequate to detect this rare threatened species (Ebner et al. 2008). On eleven and two occasions the metabarcoding analyses failed to detect species while the expert opinion and the electrofishing data indicated their presence, respectively (Suppl. material 1).

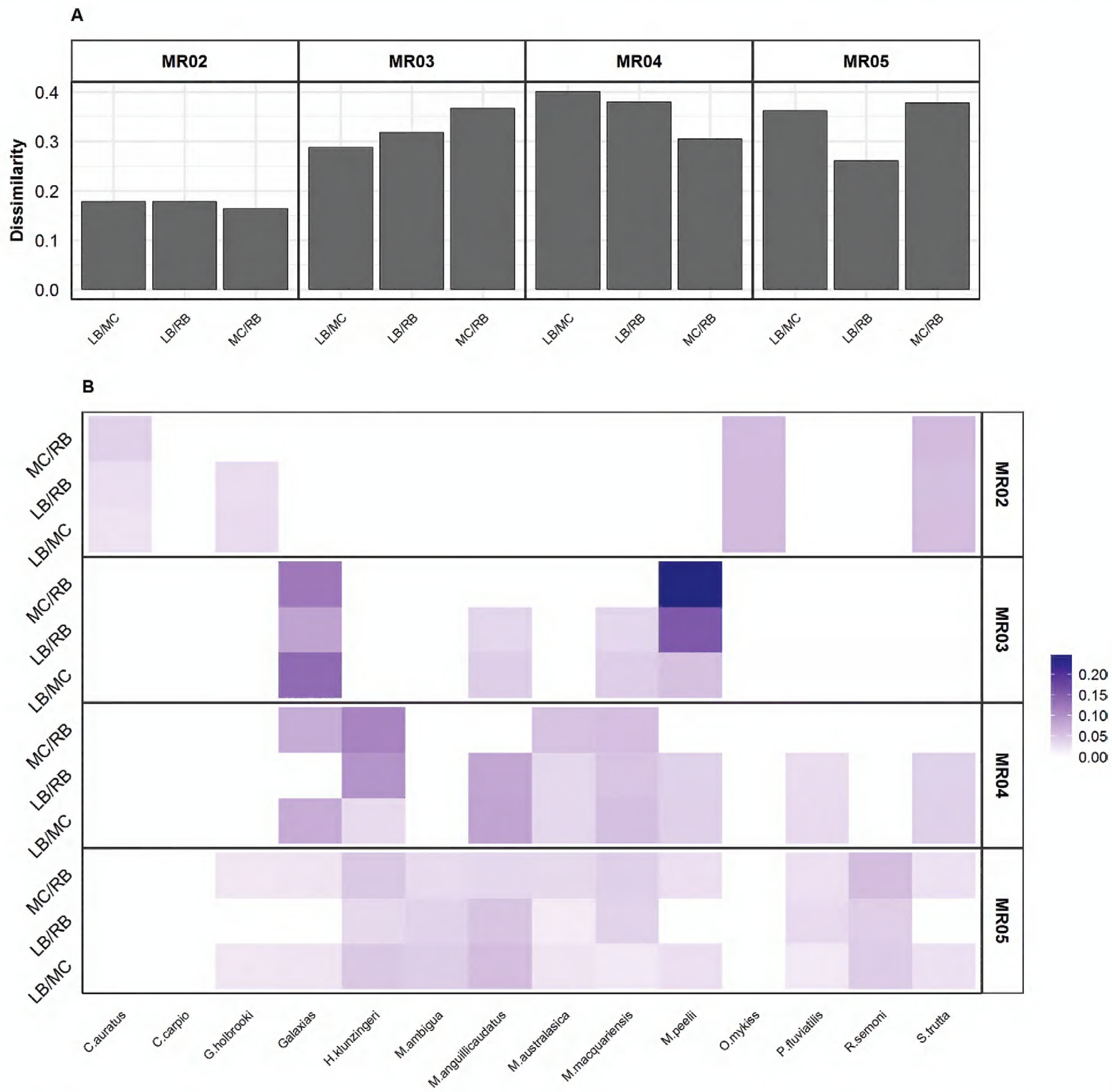

Figure 4. Overall dissimilarity between the fish community data obtained from the samples collected from the left bank (LB), mid-channel (MC) and right bank (RB) for the different sampling sites (**A**). The heat map shows the average contribution of each species to the overall community dissimilarity (**B**).

These failed detections could be due to a very low species abundances, an extremely patchy distribution (both temporal and spatial) of these species within the sites, the influence of primer biases and/or an overestimation of the species richness by the experts.

While previous studies have found that sampling replication will improve the species richness estimates of eDNA metabarcoding surveys (Shaw et al. 2016, Olds et al. 2016, Li et al. 2018, Pont et al. 2018), the current findings suggests that the sampling intensity and the spatial sampling replication in these studies was still low and may have contributed to the failure to detect some species. The current results show that the relative importance of both sampling intensity and spatial sampling replication varies along an altitudinal and biodiversity gradient. The lack of between-sample variation in the most upstream sampling location suggests that eDNA is homogeneously distributed and that relatively little replication is needed. In the mid-altitude reaches (i.e. MR02 and MR03) some spatial eDNA heterogeneity was observed. However, the collection of samples along the river-banks appears to be a suitable sampling strategy at these sites (Figure 3A). Further downstream (i.e. MR04 and MR05) the species richness increases and a sufficiently high sampling intensity combined with appropriate spatial sampling replication is needed to characterize the species richness (Figure 3A). While both ecological factors (i.e. spatial structuring of the species community) and river morphology (i.e. width, depth and water flow) may contribute to the observed results, follow up studies are needed to determine the rela-

tive importance of both biotic and abiotic factors. For example, seasonal sampling surveys can potentially reveal the influence of water flow on the spatial distribution of eDNA in riverine systems (Jane et al. 2015, Shogren et al. 2017). On the other hand, surveys in riverine systems with a very similar geology/hydrology but with different species communities and/or a different spatial arrangement of instream habitat would be valuable to determine whether eDNA metabarcoding results are able to detect the community structure at small spatial scales. In contrast to the observed effects of sampling intensity and spatial sampling replication, the impact of sequencing depth does not follow a clear pattern. The effect of sequencing depth was more profound at those sites with a low evenness in the proportional read abundance (Figure 3). However, predicting the appropriate sequencing depth is difficult, as the proportional read data from eDNA metabarcoding surveys will be affected by the species richness, the relative abundance of species, species-specific seasonal variation in eDNA concentrations and primer specific amplification biases (Elbrecht and Leese 2015, Pinol et al. 2015, Buxton et al. 2017, Bylemans et al. 2017, Takahashi et al. 2017). Consequently, pilot studies are needed to accurately assess the impact of sequencing depth within a site and/or system.

Conclusion

The current study has shown that the sampling design can have a profound effect on the performance of eDNA metabarcoding survey in riverine systems. However, more research is needed to determine the universality of the patterns described here. Additionally, temporal sampling strategies will also need to be considered in future studies as eDNA concentrations undergo seasonal fluctuations (Buxton et al. 2017, Takahashi et al. 2017). While further research may provide general guidelines for the design of optimal sampling strategies, pilot studies will remain invaluable to maximize the performance of eDNA metabarcoding surveys.

Data accessibility

The summarized data of the electrofishing, expert opinion and eDNA metabarcoding surveys are available in the Data directory of the Suppl. material 2 zip folder. Additionally, all scripts used to analyse the data are available in the Rscripts folder in the Suppl. material 2.

Acknowledgments

We wish to thank the Invasive Animals Cooperative Research Centre (grant/award number: 1.W.2) and the Holsworth Wildlife Research Endowment (grant/award number: 164) for funding this research. Finally, we wish to acknowledge Dr. Owen S. Wangensteen and Dr. Dirk Steinke for the valuable feedback on the previous version of the manuscript.

References

Alberdi A, Aizpurua O, Gilbert MTP, Bohmann K (2017) Scrutinizing key steps for reliable metabarcoding of environmental samples. Methods in Ecology and Evolution 9(1): 134–47. https://doi.org/10.1111/2041-210X.12849

Barba M De, Miquel C, Boyer F, Mercier C, Rioux D, Coissac E, Taberlet P (2014) DNA metabarcoding multiplexing and validation of data accuracy for diet assessment: application to Omnivorous diet. Molecular Ecology Resources 14(2): 306–23. https://doi.org/10.1111/1755-0998.12188

Bies JM, Fox CN, Neal JW (2016) Comparison of electrofishing and gill nets for sampling Cichlid species. North American Journal of Fisheries Management 36(5). Taylor & Francis: 975–81. https://doi.org/10.1080/02755947.2016.1184203

Bista I, Carvalho GR, Walsh K, Seymour M, Hajibabaei M, Lallias D, Christmas M, Creer S (2017) Annual time-series analysis of aqueous eDNA reveals ecologically relevant dynamics of lake ecosystem biodiversity. Nature Communications 8: 14087. https://doi.org/10.1038/ncomms14087

Blackman RC, Constable D, Hahn C, Sheard AM, Durkota J, Hänfling B, Handley LL (2017) Detection of a new non-native freshwater species by DNA metabarcoding of environmental samples – first record of *Gammarus Fossarum* in the UK. Aquatic Invasions 12(2): 177–89. https://doi.org/10.3391/ai.2017.12.2.06

Bohmann K, Evans A, Gilbert MTP, Carvalho GR, Creer S, Knapp M, Yu DW, Bruyn M de (2014) Environmental DNA for wildlife biology and biodiversity monitoring. Trends in Ecology & Evolution 29 (6): 358–67. https://doi.org/10.1016/j.tree.2014.04.003

Bolger AM, Lohse M, Usadel B (2014) Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics 30(15): 2114–20. https://doi.org/10.1093/bioinformatics/btu170

Bower DS, Pickett EJ, Stockwell MP, Pollard CJ, Garnham JI, Sanders MR, Clulow J, Mahony MJ (2014) Evaluating monitoring methods to guide adaptive management of a threatened Amphibian (*Litoria Aurea*). Ecology and Evolution 4(8): 1361–68.

Boyer F, Mercier C, Bonin A, Bras Y Le, Taberlet P, Coissac E (2016) Obitools: A Unix-inspired software package for DNA metabarcoding. Molecular Ecology Resources 16(1): 176–82. https://doi.org/10.1111/1755-0998.12428.

Buxton AS, Groombridge JJ, Zakaria NB, Griffiths RA (2017) Seasonal variation in environmental DNA in relation to population size and environmental factors. Scientific Reports 7. Nature Publishing Group: 46294. https://doi.org/10.1038/srep46294.

Bylemans J, Furlan EM, Hardy CM, McGuffie P, Lintermans M, Gleeson DM (2017) An environmental DNA (eDNA) based method for monitoring spawning activity: a case study, using the endangered Macquarie Perch (*Macquaria Australasica*). Methods in Ecology and Evolution 8(5): 646–55. https://doi.org/10.1111/2041-210X.12709

Bylemans J, Gleeson DM, Hardy CM, Furlan E (2018) Toward an ecoregion scale evaluation of eDNA metabarcoding primers: a case

study for the freshwater fish biodiversity of the Murray-Darling basin (Australia). Ecology and Evolution 8(17): 8697–8712. https://doi.org/10.1002/ece3.4387

Clavero M, Blanco-Garrido F, Prenda J (2006) Monitoring small fish populations in streams: a comparison of four passive methods. Fisheries Research 78(2–3): 243–51. https://doi.org/10.1016/j.fishres.2005.11.016

Creer S, Deiner K, Frey S, Porazinska D, Taberlet P, Thomas w. K, Potter C, Bik HM (2016) The ecologist's field guide to sequence-based identification of biodiversity. Methods in Ecology and Evolution 7(9): 1008–18. https://doi.org/10.1111/2041-210X.12574

Deiner K, Bik HM, Mächler E, Seymour M, Lacoursière-Roussel A, Altermatt F, Creer S et al. (2017) Environmental DNA metabarcoding: transforming how we survey animal and plant communities. Molecular Ecology 26(21): 5872–95. https://doi.org/10.1111/mec.14350

Deiner K, Walser J-C, Mächler E, Altermatt F (2015) Choice of capture and extraction methods affect detection of freshwater biodiversity from environmental DNA. Biological Conservation 183: 53–63. https://doi.org/10.1016/j.biocon.2014.11.018

Dickie IA, Boyer S, Buckley H, Duncan RP, Gardner P, Hogg ID, Holdaway RJ et al. (2018) Towards robust and repeatable sampling methods in eDNA based studies. Molecular Ecology Resources 18(5): 940–52. https://doi.org/10.1111/1755-0998.12907

Ebner BC, Thiem JD, Gilligan DM, Lintermans M, Wooden IJ, Linke S (2008) Estimating species richness and catch per unit effort from boat electro-fishing in a Lowland river in Temperate Australia. Austral Ecology 33(7): 891–901. https://doi.org/10.1111/j.1442-9993.2008.01862.x

Elbrecht V, Leese F (2015) Can DNA-based ecosystem assessments quantify species abundance? Testing primer bias and biomass-sequence relationships with an innovative metabarcoding protocol. PLoS ONE 10(7): 1–16. https://doi.org/10.1371/journal.pone.0130324

Elbrecht V, Leese F (2017) Validation and development of COI metabarcoding primers for freshwater macroinvertebrate bioassessment. Frontiers in Environmental Science 5: 11. https://doi.org/10.3389/fenvs.2017.00011

Evans NT, Li Y, Renshaw MA, Olds BP, Deiner K, Turner CR, Jerde CL et al. (2017) Fish community assessment with eDNA metabarcoding: effects of sampling design and bioinformatic filtering. Canadian Journal of Fisheries and Aquatic Sciences 74(9): 1362–74. https://doi.org/10.1139/cjfas-2016-0306

Faircloth BC, Glenn TC (2012) Not all sequence tags are created equal: designing and validating sequence identification tags robust to indels. PLoS ONE 7(8): e42543. https://doi.org/10.1371/journal.pone.0042543

Ficetola GF, Miaud C, Pompanon F, Taberlet P (2008) Species detection using environmental DNA from water samples. Biology Letters 4(4): 423–25. https://doi.org/10.1098/rsbl.2008.0118

Ficetola GF, Pansu J, Bonin A, Coissac E, Giguet-Covex C, Barba M De, Gielly L et al. (2015) Replication levels, false presences, and the estimation of presence / absence from eDNA metabarcoding data. Molecular Ecology Resources 15(3): 543–56. https://doi.org/10.1111/1755-0998.12338

Ficetola GF, Taberlet P, Coissac E (2016) How to limit false positives in environmental DNA, metabarcoding? Molecular Ecology Resources 16(3): 604–7. https://doi.org/10.1111/1755-0998.12508

Fischer JR, Quist MC (2014) Characterizing Lentic freshwater fish assemblages using multiple sampling methods. Environmental Monitoring and Assessment 186(7): 4461–74. https://doi.org/10.1007/s10661-014-3711-z

Goldberg CS, Pilliod DS, Arkle RS, Waits LP (2011) Molecular detection of vertebrates in stream water: a demonstration using Rocky Mountain tailed frogs and Idaho giant salamanders. PloS One 6(7): e22746. https://doi.org/10.1371/journal.pone.0022746

Gotelli NJ, Colwell RK (2001) Quantifying biodiversity: procedures and pitfalls in the measurement and comparison of species richness. Ecology Letters 4(4): 379–91. https://doi.org/10.1046/j.1461-0248.2001.00230.x

Hänfling B, Handley LL, Read DS, Hahn C, Li J (2016) Environmental DNA metabarcoding of lake fish communities reflects long-term data from established survey methods. Molecular Ecology 25(13): 3101–3119. https://doi.org/10.1111/mec.13660

Hinlo R, Lintermans M, Gleeson D, Broadhurst B, Furlan E (2018) Performance of eDNA assays to detect and quantify an elusive benthic fish in upland streams. Biological Invasions. Springer International Publishing. https://doi.org/10.1007/s10530-018-1760-x

Hsieh TC, Ma KH, Chao A (2016) INEXT: An R package for rarefaction and extrapolation of species diversity (hill numbers). Methods in Ecology and Evolution 7(12): 1451–56. https://doi.org/10.1111/2041-210X.12613

Jane SF, Wilcox TM, McKelvey KS, Young MK, Schwartz MK, Lowe WH, Letcher BH, Whiteley AR (2015) Distance, flow, and PCR inhibition: eDNA dynamics in two headwater streams. Molecular Ecology Resources 15(1): 216–27. https://doi.org/10.1111/1755-0998.12285

Jerde CL, Mahon AR, Chadderton WL, Lodge DM (2011) "Sight-Unseen" detection of rare aquatic species using environmental DNA. Conservation Letters 4(2): 150–57. https://doi.org/10.1111/j.1755-263X.2010.00158.x

Jurajda P, Janáč M, White SM, Ondračková M (2009) Small – but not easy: evaluation of sampling methods in floodplain lakes including Whole-Lake sampling. Fisheries Research 96(1): 102–8. https://doi.org/10.1016/j.fishres.2008.09.005

Kanagawa T (2003) Bias and artifacts in multitemplate polymerase chain reactions (PCR). Journal of Bioscience and Bioengineering 96(4): 317–23. https://doi.org/10.1016/S1389-1723(03)90130-7

Kennard MJ, Pusey BJ, Harch BD, Dore E, Arthington AH (2006) Estimating local stream fish assemblage attributes: sampling effort and efficiency at two spatial scales. Marine and Freshwater Research 57(6): 635–53. https://doi.org/10.1071/MF06062

Li Y, Evans NT, Renshaw MA, Jerde CL, Olds BP, Shogren AJ, Deiner K et al. (2018) Estimating fish Alpha- and Beta-diversity along a small stream with environmental DNA metabarcoding. Metabarcoding and Metagenomics 2: e24262. https://doi.org/10.3897/mbmg.2.24262

Lintermans M (2002) Fish in the Upper Murrumbidgee Catchment: A review of current knowledge. Canberra, Australia. https://upperbidgeereach.org.au/files/Lintermans-2002FishUpperbidgee.pdf

Lintermans M (2007) Fishes of the Murray-Darling Basin: An Introductory Guide. https://www.mdba.gov.au/sites/default/files/pubs/MDBA-Fish-species-book.pdf

Lintermans M (2016) Finding the needle in the haystack: comparing sampling methods for detecting an endangered freshwater fish. Marine and Freshwater Research 67: 1740–49. https://doi.org/10.1071/MF14346

Miya M, Sato Y, Fukunaga T, Sado T, Poulsen JY, Sato K, Minamoto T et al. (2015) MiFish , a set of universal PCR primers for metabarcoding environmental dna from fishes: detection of more than 230 subtropical marine species. Royal Society Open Science 2: 150088. https://doi.org/10.1098/rsos.150088

O'Donnell JL, Kelly RP, Lowell NC, Port JA (2016) Indexed PCR primers induce template-specific bias in large-scale DNA sequencing studies. Plos One 11(3): e0148698. https://doi.org/10.1371/journal.pone.0148698

Oberdorff T, Guilbert E, Lucchetta J-C (1993) Patterns of fish species richness in the Seine river basin, France. Hydrobiologia 259(3): 157–67. https://doi.org/10.1007/BF00006595

Oksanen J, Kindt R, Legendre P, O'Hara B, Stevens MHH, Oksanen MJ, Suggests M (2007) The vegan package. Community Ecology Package 10: 631–37.

Olds BP, Jerde CL, Renshaw MA, Li Y, Evans NT, Turner CR, Deiner K et al. (2016) Estimating species richness using environmental DNA. Ecology and Evolution 6(12): 4214–26. https://doi.org/10.1002/ece3.2186

Pinol J, Mir G, Gomez-Polo P, Agusti N (2015) Universal and blocking primer mismatches limit the use of high-throughput DNA sequencing for the quantitative metabarcoding of Arthropods. Molecular Ecology Resources 15(4): 819–30. https://doi.org/10.1111/1755-0998.12355

Platts WS (1979) Relationships among stream order, fish populations, and aquatic geomorphology in an Idaho River Drainage. Fisheries 4(2): 5–9. https://doi.org/10.1577/1548-8446(1979)004<0005:RASOFP>2.0.CO;2

Pont D, Rocle M, Valentini A, Civade R, Jean P, Maire A, Roset N et al. (2018) Environmental DNA reveals quantitative patterns of fish biodiversity in large rivers despite its downstream transportation. Scientific Reports 8(1): 1–13. https://doi.org/10.1038/s41598-018-28424-8

Porreca AP, Pederson CL, Laursen JR, Colombo RE (2013) A comparison of electrofishing methods and fyke netting to produce reliable abundance and size metrics. Journal of Freshwater Ecology 28(4): 585–90. https://doi.org/10.1080/02705060.2013.810555

Port JA, O'Donnell JL, Romero-Maraccini OC, Leary PR, Litvin SY, Nickols KJ, Yamahara KM, Kelly RP (2016) Assessing vertebrate biodiversity in a Kelp forest ecosystem using environmental DNA. Molecular Ecology 25(2): 527–41. https://doi.org/10.1111/mec.13481

Pritt JJ, Frimpong EA (2014) The effect of sampling intensity on patterns of rarity and community assessment metrics in stream fish samples. Ecological Indicators 39: 169–78. https://doi.org/10.1016/j.ecolind.2013.12.016

R Development Core Team (2010) R: A language and environment for statistical computing. R Foundation for Statistical Computing Vienna (Austria): ISBN 3-900051-07-0.

Rees HC, Maddison BC, Middleditch DJ, Patmore JRM, Gough KC (2014) The detection of aquatic animal species using environmental DNA - a review of eDNA as a survey tool in ecology. Journal of Applied Ecology 51(5): 1450–59. https://doi.org/10.1111/1365-2664.12306

Reid SM, Haxton TJ (2017) Backpack electrofishing effort and imperfect detection: influence on riverine fish inventories and monitoring. Journal of Applied Ichthyology 33(6): 1083–91. https://doi.org/10.1111/jai.13463

Ruetz CR, Uzarski DG, Krueger DM, Rutherford ES (2007) Sampling a littoral fish assemblage: comparison of small-mesh fyke netting and boat electrofishing. North American Journal of Fisheries Management 27(3): 825–31. https://doi.org/10.1577/M06-147.1

Sato H, Sogo Y, Doi H, Yamanaka H (2017) Usefulness and limitations of sample pooling for environmental DNA metabarcoding of freshwater fish communities. Scientific Reports 7: 14860. https://doi.org/10.1038/s41598-017-14978-6

Schmidt BR (2005) Monitoring the distribution of pond-breeding Amphibians when species are detected imperfectly. Aquatic Conservation: Marine and Freshwater Ecosystems 15(6): 681–92. https://doi.org/10.1002/aqc.740

Schnell IB, Bohmann K, Gilbert MTP (2015) Tag jumps illuminated - reducing sequence-to-sample misidentifications in metabarcoding studies. Molecular Ecology Resources 15(6): 1289–1303. https://doi.org/10.1111/1755-0998.12402

Shaw JLA, Clarke LJ, Wedderburn SD, Barnes TC, Weyrich LS, Cooper A (2016) Comparison of environmental DNA metabarcoding and conventional fish survey methods in a river system. Biological Conservation 197: 131–38. https://doi.org/10.1016/j.biocon.2016.03.010

Shelton AO, O'Donnell JL, Samhouri JF, Lowell N, Williams GD, Kelly RP (2016) A framework for inferring biological communities from environmental DNA. Ecological Applications 26(6): 1645–59. https://doi.org/10.1890/15-1733.1

Shogren AJ, Tank JL, Andruszkiewicz E, Olds B, Mahon AR, Jerde CL, Bolster D (2017) Controls on eDNA movement in streams: transport, retention, and resuspension. Scientific Reports 7(1). Springer US: 5065. https://doi.org/10.1038/s41598-017-05223-1

Sigsgaard EE, Carl H, Møller PR, Thomsen PF (2015) Monitoring the near-extinct European weather Loach in Denmark based on environmental DNA from water samples. Biological Conservation 183: 46–52. https://doi.org/10.1016/j.biocon.2014.11.023

Smith TA, Kraft CE (2005) Stream fish assemblages in relation to landscape position and local habitat variables. Transactions of the American Fisheries Society 134(2): 430–40. https://doi.org/10.1577/T03-051.1

Stoeckle MY, Soboleva L, Charlop-Powers Z (2017) Aquatic environmental DNA detects seasonal fish abundance and habitat preference in an urban estuary. PLoS ONE 12(4): e0175186. https://doi.org/10.1371/journal.pone.0175186

Taberlet P, Coissac E, Pompanon F, Brochmann C, Willerslev E (2012) Towards next-generation biodiversity assessment using DNA metabarcoding. Molecular Ecology 21(8): 2045–50. https://doi.org/10.1111/j.1365-294X.2012.05470.x

Takahashi MK, Meyer MJ, Mcphee C, Gaston JR, Venesky MD, Case BF (2017) Seasonal and diel signature of eastern hellbender environmental DNA. The Journal of Wildlife Management 82: 217–25. https://doi.org/10.1002/jwmg.21349

Tejerina-Garro FL, Maldonado M, Ibanez C, Pont D, Roset N, Oberdorff T (2005) Effects of natural and anthropogenic environmental changes on riverine fish assemblages: a framework for ecological assessment of rivers. Brazilian Archives of Biology and Technology 48: 91–108.

Unmack PJ (2001) Biogeography of Australian freshwater fishes. Journal of Biogeography 28: 1053–89.

Valentini A, Taberlet P, Miaud C, Civade R, Herder J, Thomsen PF, Bellemain E et al. (2016) Next-generation monitoring of aquatic biodiversity using environmental DNA metabarcoding. Molecular Ecology 25(4): 929–42. https://doi.org/10.1111/mec.13428

Wickham H (2016) Tidyverse: easily install and load "Tidyverse" packages. R Package.

Yamamoto S, Masuda R, Sato Y, Sado T, Araki H, Kondoh M, Minamoto T, Miya M (2017) Environmental DNA metabarcoding reveals local fish communities in a species-rich coastal sea. Scientific Reports 7: 40368. https://doi.org/10.1038/srep40368

Supplementary material 1
Method overview and data summaries
Authors: Jonas Bylemans, Dianne M. Gleeson, Mark Lintermans, Christopher M. Hardy, Matthew Beitzel, Dean M. Gilligan, Elise M. Furlan
Data type: Microsoft Word Document (.docx)
Copyright notice: This dataset is made available under the Open Database License (http://opendatacommons.org/licenses/odbl/1.0/). The Open Database License (ODbL) is a license agreement intended to allow users to freely share, modify, and use this Dataset while maintaining this same freedom for others, provided that the original source and author(s) are credited.
Link: https://doi.org/10.3897/mbmg.2.30457.suppl1

Supplementary material 2
Data and R script
Authors: Jonas Bylemans, Dianne M. Gleeson, Mark Lintermans, Christopher M. Hardy, Matthew Beitzel, Dean M. Gilligan, Elise M. Furlan
Data type: statistical data
Copyright notice: This dataset is made available under the Open Database License (http://opendatacommons.org/licenses/odbl/1.0/). The Open Database License (ODbL) is a license agreement intended to allow users to freely share, modify, and use this Dataset while maintaining this same freedom for others, provided that the original source and author(s) are credited.
Link: https://doi.org/10.3897/mbmg.2.30457.suppl2